Not by chance. Russian aspect in rule-based machine translation

Sonnenhauser, Barbara ; Zangenfeind, Robert

Abstract: The aim of this paper is twofold: it illustrates the benefits of rule-based instead of statistical machine translation, and it provides a starting point for the machine translation of the Russian aspect into English. Rule-based machine translation is still promising, from both a computational and theoretical point of view, because by implementing rules on the computer theoretical assumptions concerning linguistic structures can be verified and improved. This will be shown using the example of the category of aspect, which is one of the main challenges for machine translation from Russian to English. A small corpus study on the translation of Russian sentences with verbs in the past tense (perfective and imperfective) by human translators shows that three-quarters of Russian verbs (both imperfective and perfective) are translated by English simple past forms. While this results from language internal markedness relations, the translation of the remaining 25 % requires an in-depth analysis of the various interpretations possible for the Russian aspect. We propose a semantic analysis based on which rules for the interpretation and translation of Russian aspect in a machine translation system can be derived. Their implementation in the machine translation system ETAP is shown in this paper using two test cases as examples.

DOI: https://doi.org/10.1007/s11185-016-9169-6

Metadata of the article that will be visualized in Online First

Journal Name	Russian Linguistics	
Article Title	Not by chance. Russian aspect in rule-based machine translation	
Copyright holder	Springer Science+Business Media Dordrecht This will be the copyright line in the final PDF.	
Corresponding Author	Family name	Sonnenhauser
	Particle	
	Given Name	Barbara
	Suffix	
	Division	Slavisches Seminar
	Organization	Universität Zürich
	Address	Zurich, Switzerland
	E-mail	barbara.sonnenhauser@uzh.ch
Author	Family name	Zangenfeind
	Particle	
	Given Name	Robert
	Suffix	
	Division	Institut für Slavische Philologie
	Organization	Ludwig-Maximilians-Universität München
	Address	Munich, Germany
	E-mail	robert.zangenfeind@slav.fak12.uni-muenchen.de
Schedule	Received	
	Revised	
	Accepted	

Abstract

The aim of this paper is twofold: it illustrates the benefits of rule-based instead of statistical machine translation, and it provides a starting point for the machine translation of the Russian aspect into English. Rule-based machine translation is still promising, from both a computational and theoretical point of view, because by implementing rules on the computer theoretical assumptions concerning linguistic structures can be verified and improved. This will be shown using the example of the category of aspect, which is one of the main challenges for machine translation from Russian to English. A small corpus study on the translation of Russian sentences with verbs in the past tense (perfective and imperfective) by human translators shows that three-quarters of Russian verbs (both imperfective and perfective) are translated by English simple past forms. While this results from language internal markedness relations, the translation of the remaining 25 % requires an in-depth analysis of the

	various interpretations possible for the Russian aspect. We propose a semantic analysis based on which rules for the interpretation and translation of Russian aspect in a machine translation system can be derived. Their implementation in the machine translation system ĖTAP is shown in this paper using two test cases as examples.
Аннотация	Цель этой статьи двояка: она иллюстрирует пользу машинного перевода на основе правил по сравнению с машинным переводом на основе статистики и предлагает отправной пункт для машинного перевода русского вида глаго ла на английский язык. Машинный перевод на основе правил всё ещё име ет свои выгоды, и с вычислительной, и с теоретической точки зрения, поско льку, применив правила на компьютере, теоретические гипотезы, касающие ся лингвистических структур, будут проверены и улучшены. Мы это покаже м на примере вида глагола, который является одной из главных сложносте й для машинного перевода с русского на английский язык. Исследуя часть параллельного корпуса русского национального корпуса, мы изучаем, как ру сские предложения с глаголами в прошедшем времени переводятся на анг лийский язык переводчиками-людьми. Эти исследования показывают, что тр и четверти русских глаголов (как несовершенного, так и совершенного вида) этого корпуса переводятся английскими формами *past simple* (претерит). В то время как это представляет собой следствие внутренних языковых от ношений маркированности, перевод остальных 25 % требует глубокого анал иза различных возможностей интерпретации русского аспекта. На основе се мантического анализа, который мы предложим, можно получить правила дл я трактовки и перевода русского аспекта в системе машинного перевода. И х применение в системе машинного перевода (в этом случае ЭТАП) продем онстрирована в данной статье на двух примерах.
Keywords	
Footnotes	

Russ Linguist
DOI 10.1007/s11185-016-9169-6

Not by chance. Russian aspect in rule-based machine translation

Не случайно. Машинный перевод русского глагольного вида на основе правил

Barbara Sonnenhauser[1] · **Robert Zangenfeind**[2]

Abstract The aim of this paper is twofold: it illustrates the benefits of rule-based instead of statistical machine translation, and it provides a starting point for the machine translation of the Russian aspect into English. Rule-based machine translation is still promising, from both a computational and theoretical point of view, because by implementing rules on the computer theoretical assumptions concerning linguistic structures can be verified and improved. This will be shown using the example of the category of aspect, which is one of the main challenges for machine translation from Russian to English. A small corpus study on the translation of Russian sentences with verbs in the past tense (perfective and imperfective) by human translators shows that three-quarters of Russian verbs (both imperfective and perfective) are translated by English simple past forms. While this results from language internal markedness relations, the translation of the remaining 25 % requires an in-depth analysis of the various interpretations possible for the Russian aspect. We propose a semantic analysis based on which rules for the interpretation and translation of Russian aspect in a machine translation system can be derived. Their implementation in the machine translation system ÉTAP is shown in this paper using two test cases as examples.

Аннотация Цель этой статьи двояка: она иллюстрирует пользу машинного перевода на основе правил по сравнению с машинным переводом на основе статистики и предлагает отправной пункт для машинного перевода русского вида глагола на английский язык. Машинный перевод на основе правил всё ещё имеет свои выгоды, и с вычислительной, и с теоретической точки зрения, поскольку, применив правила на компьютере, теоретические гипотезы, касающиеся лингвистических структур, будут проверены и улучшены. Мы это покажем на примере вида глагола, который является одной из главных сложностей для машинного перевода с русского на английский язык. Исследуя часть параллельного корпуса русского национального корпуса, мы

✉ B. Sonnenhauser
barbara.sonnenhauser@uzh.ch

R. Zangenfeind
robert.zangenfeind@slav.fak12.uni-muenchen.de

[1] Slavisches Seminar, Universität Zürich, Zurich, Switzerland

[2] Institut für Slavische Philologie, Ludwig-Maximilians-Universität München, Munich, Germany

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 2/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

изучаем, как русские предложения с глаголами в прошедшем времени переводятся на английский язык переводчиками-людьми. Эти исследования показывают, что три четверти русских глаголов (как несовершенного, так и совершенного вида) этого корпуса переводятся английскими формами *past simple* (претерит). В то время как это представляет собой следствие внутренних языковых отношений маркированности, перевод остальных 25 % требует глубокого анализа различных возможностей интерпретации русского аспекта. На основе семантического анализа, который мы предложим, можно получить правила для трактовки и перевода русского аспекта в системе машинного перевода. Их применение в системе машинного перевода (в этом случае ЭТАП) продемонстрировано в данной статье на двух примерах.

1 Introduction

In the last decades, statistical approaches have been gaining more and more ground in machine translation. This has been facilitated by the increasing availability of large-scale parallel corpora. However, there are two main drawbacks to this approach: First, corpora of the necessary size and quality are not available for all pairs of languages, and second, a superficially high success rate of statistical machine translation may rely on pure chance, i.e. on simple 'default mappings'. The first caveat holds for Russian-English translation in general, the second for the translation of morphological categories that do not have exact counterparts in both languages, such as the grammatical category of aspect. Using this particular problem as starting point, the present paper intends to show that a rule-based approach is—despite all progress of statistical machine translation—still promising, from both a computational and theoretical point of view.

The paper is structured as follows: Sect. 2 illustrates some benefits of rule-based machine translation in general. Based on examples taken from professional translations of 20th century Russian literature, Sect. 3 presents a case study that demonstrates some of the problems involved in the analysis and interpretation of the Russian imperfective and perfective aspect, and illustrates the ways in which these problems are relevant for machine translation. In Sect. 4 the implementation of some first rules for the interpretation and translation of Russian aspect in a machine translation system is described. The main findings are summarised in Sect. 5.

2 The usefulness of rule-based machine translation

On the one hand, statistical machine translation (SMT) has developed significantly over the last years. This is not only because electronic corpora are available in growing numbers but also because the algorithms for translation (i.e. pure translation of words or short sequences of words) and for adjusting word order (in the target language) have been refined. As a consequence, SMT yields passably good results for particular language pairs and specific types of texts.[1]

[1] Graham, Baldwin, Moffat, and Zobel (2014) examined the quality of about 40 MT systems for seven language pairs. For the translation of news articles from English to Spanish they note the following results for the best SMT system: fluency of the translated texts was evaluated as almost 72 % by human evaluators (rating the text by how much they agreed that "the text is fluent Spanish" (ibid., p. 445) on a 100-point scale); adequacy was evaluated as about 67 % (rating the text by how much the human evaluators agreed that it "adequately

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 3/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

On the other hand, rule-based machine translation (RBMT) is still beneficial for linguistics. Since for RBMT dictionaries and grammatical rules are necessary that represent linguistic competence, it can enhance knowledge about natural languages as concerns, e.g., lexical semantics and morphosyntax. For SMT, on the contrary, no explicit knowledge of the particular languages is necessary; good translations are produced when the data taken from already existing (bilingual and monolingual) large corpora are computed by good algorithms. That is, one advantage of applying RBMT instead of a pure statistical approach is that it is a valuable source for theoretical linguistics (cf. Iomdin 2003, 2008; Apresjan et al. 1989).

Apresjan et al. (1989, p. 285) list five domains in which theoretical linguistics may profit from RBMT: 1) the description of particular languages, 2) the modelling of theoretical principles, 3) the development of a meta-language, 4) questions of contrastive linguistics, and 5) problems related to the comprehension of texts. Regarding the MT of grammatical categories, domains 2 and 3 are of particular relevance, i.e. the modelling of rules derived from theoretical principles and based on an appropriate meta-linguistic description of the semantic and syntactic features figuring in those principles. If these principles and the corresponding rules are implemented in an appropriate way, the computer will produce correct translations. If it does not, it is obvious that the underlying rules have to be reformulated or new rules need to be added to the system. Thus, theoretical assumptions can be verified and improved in practice, which in turn contributes to the shaping of linguistic description. Apresjan et al. (1989, p. 285) point out that errors produced by the computer differ from those made by a human translator (cf. example (1) below). Thereby, computer generated mistakes provide 'unique negative linguistic material' ("unikal'nyj 'otricatel'nyj' jazykovoj material", ibid.) in that these mistakes direct attention to linguistic features that are otherwise hardly identifiable. In this respect, insight gained from errors occurring in the course of RBMT may even be compared to the significance of aphasia for the understanding of the mechanisms of thought (ibid.).

How incorrect translations of a RBMT system can help to improve linguistic theory is illustrated by Iomdin (2003) based on the example of the erroneous automatic parsing of the Russian sentence in (1), which resulted in a wrong translation into English. In order to fix this wrong interpretation, the syntactic structure was analysed. This revealed a particular syntactic property which characterises a group of Russian nouns (*ideja* 'idea', *tezis* 'thesis', *rezul'tat* 'result' etc.), but not another group with a similar semantics (*cel'* 'purpose', *želanie* 'wish', *problema* 'problem' etc.) and, thus, makes a specific syntactic construction possible for the first group of nouns but not for the second group, as illustrated in (1):

(1) Osnovnaja *ideja* [**cel'*] konkursa—pust' pobedit sil'nejšij.
'The main *idea* of the competition is: let the strongest win.' (Iomdin 2003, p. 256)

Nouns of the type *ideja* 'idea' have the capability to serve as a syntactic subject in copulative sentences in which the complement of the copula, i.e. the predicative, is a predicative clause, as in (1). Nouns of the type '*cel'*', on the other hand, have the capability to serve as a syntactic subject in copulative sentences in which the complement of the copula is an infinitive.[2] This

expresses the meaning" (ibid., p. 447) of the translation by a human translator). The improvement between 2007 and 2012 amounted to about 12 percentage points. The difficulties involved in measuring the quality of the results of machine translation, in particular ÉTAP, are pointed out by Apresjan et al. (1989, p. 11). Referring to Kulagina (1979), Apresjan et al. (1989, p. 11) cite the following three criteria—similar to 'fluency' and 'adequacy' listed by Graham et al. (2014), but different in detail—, that all have to be evaluated by an expert (in the end: a human): 1) degree of match in terms of content, 2) comprehensibility, 3) grammatical correctness. While criterion 3) seems to be measurable on largely objective grounds, 1) and 2) crucially rely on native speakers' intuitions.

[2]Cf. "Naša cel'—ustanovit' istinu 'Our purpose is to establish the truth' " (Iomdin 2003, p. 255).

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 4/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

is described by the syntactic feature 'predinf' (meaning 'the predicative can be an infinitive') for these nouns in their dictionary entries.

In order to capture the difference between the *ideja*-type and the *cel'*-type, a new syntactic feature had to be introduced for the former: 'predsent' (meaning 'the predicative can be a sentence'). By introducing this syntactic feature the parser could be fixed and the sentence in (1) was translated correctly.

By introducing a specific syntactic feature ('predsent', by analogy with 'predinf' which characterises the '*cel'*-group') for the according lexemes the parser could be fixed and the sentence was translated correctly.

3 Translating Russian verbal aspect

Among the main challenges for machine translation from Russian is the category of aspect. The Russian imperfective (ipf) and perfective (pf) aspects cause considerable problems for machine translation, since they give rise to a range of potential interpretations that may require the usage of different translation equivalents. The difficulties for cross-linguistic comparison relating to the manifold readings for both aspects have been extensively discussed in literature (cf., e.g., Sonnenhauser 2006 for an overview). The analysis proposed in Sonnenhauser (2006, pp. 259–264) suggests that for MT, the appropriate reading has to be determined in a first step in order to provide the basis for the translation process. In this respect, RBMT helps to systematize these interpretations since it requires the formulation of appropriate disambiguation rules which can be implemented and verified at the computer. The linguistic basis for the formulation of these rules will be illustrated in this section.

3.1 Semantics and interpretation

Even though English has a morphological category of aspect, it is not 1:1-equivalent of the Russian aspect category.[3] As a consequence, English may require different morphological forms, depending on the specific function of the Russian ipf or pf aspect. This does not seem to be a big deal for human translation. As Padučeva (1992, p. 113) points out for the ipf aspect, "[t]he translation of the forms bearing the progressive or the habitual meaning into European languages poses no problem". The former can be translated into English with the progressive form, the latter "by unmarked tense forms" (ibid.), i.e. the simple past or simple present. What is problematic for translation is the 'factual meaning', as Padučeva (1992) calls it, since it covers various meaning components. They come to the fore in different contexts, "which requires different [translation] equivalents in different cases" (ibid., p. 125). What all three meanings mentioned by Padučeva have in common is that they equally require one additional step preceding the translation process: the language-internal determination of the relevant interpretation (here: progressive, habitual, factual). Only then is it possible to choose the appropriate equivalent in the target language. As a consequence, an immediate morpheme-based translation such as 'ipf → *progressive form*' 'pf → *simple form*' would

[3] Apresjan et al. (1989, p. 11) point out the difficulties related to aspect in MT from English to Russian. These problems result mainly from the fact that it is not possible to 'synthesise' the adequate Russian aspect form based on the verbal form in the English original, since English—according to Apresjan et al.—does not have an aspect category. Even though this lack of an aspect category is disputed in the present paper, Apresjan et al. are right in stating that there is no 1:1-correspondence between the English and the Russian system. That is, lexical semantic features of the verbal forms as well as context have to be taken into account. The same holds for translations from Russian into English, the topic of the present paper.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 5/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

Table 1 Groups of Russian aspect interpretation

Semantic group	Russian: interpretation	English: morphological representation
Group I_{ipf}	processual, conative	past progressive
Group II_{ipf}	habitual, non-actual, potential, permanent, atemporal	simple past
Group III_{ipf}	general-factive, durative	simple past
Group I_{pf}	eventive	simple past
Group II_{pf}	perfect (existential, current relevance, extended now, etc.)	perfect (present perfect)
Group III_{pf}	pluperfect	pluperfect (past perfect)

be inadequate for MT; however, translating context-determined interpretations would not be manageable. Both facts provide additional arguments for a rule-based system.

Judging from the above considerations, RBMT of aspect must be carried out in two steps: language-internal disambiguation towards the relevant aspect interpretation, and translation itself. The first step involves the distinction of semantically coded information, which serves as the basis for the process of translation, from contextually determined readings, which are a language-internal problem.

In Sonnenhauser and Zangenfeind (2013), and Zangenfeind and Sonnenhauser (2014) the language-internal issue of aspect interpretation has been discussed. In this first approach six semantically grounded groups of possible interpretations for Russian verbs in the past tense—three groups each for the two aspects—have been stated, together with their morphological representation in English, cf. Table 1.

The distinction of the groups listed in Table 1 is based on a time-relational and selection-theoretic approach to aspect (elaborated in Sonnenhauser 2006 on the basis of Klein 1995 and Bickel 1996), according to which the semantics of aspect can be described as the selection and assertion of some specific part of the event structure coded by the verbal predicate. The three groups of the ipf aspect are characterised by different relations between the topic time interval (i.e. the time interval an assertion is about) and the event time interval (i.e. the part of the run time of the denoted event that is selected by the aspect operator). As concerns the ipf aspect, both intervals may be congruent (group II_{ipf}), the event time interval as a whole may be included in the topic time interval (group III_{ipf}) or the topic time may be included in the process or state component coded by the verbal predicate (group I_{ipf}). The groups of the pf aspect all include an event boundary; they differ as concerns the closedness vs. openness of the boundaries of topic time interval: both are closed for group I_{pf}, the right boundary is open for group II_{pf}, the left boundary is open for group III_{pf}. The details are described in Sonnenhauser and Zangenfeind (2013).

As can be seen from Table 1, English uses different morphological means for the various interpretations of the Russian aspects. Importantly, these morphological expressions cover the same interpretational range as the three semantically distinguished groups for each aspect. Relevant for MT is thus the differentiation of these groups, while the various interpretations within each group are a language-internal problem—both for the source and target language—and do not concern the translation process as such.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 6/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

3.2 A case study

To obtain a more precise overview of the English morphological equivalents of the Russian aspect forms, we examined in a small corpus study how Russian sentences with verbs in the past tense are translated into English by human translators. These translations at the same time serve as 'expert translations' for evaluating the quality of MT (see footnote 1). The corpus was compiled with examples taken from the Russian-English parallel corpus of the Russian National Corpus (RNC),[4] which includes mainly classical literature written by Russian and English authors, respectively, and their aligned translations into the other language (cf. Dobrovol'skij, Kretov, and Šarov 2005). To have a workable basis we examined only Russian literature from the 20th century and its English translations. The corpus for our study consists of the first some twenty sentences in the chosen authors' works (A. N. and B. N. Strugackij, N. N. Nosov, M. A. Bulgakov, N. A. Ostrovskij, I. A. Il'f and E. P. Petrov, M. Gorkij, L. N. Tolstoj) in order to have a sample of different styles by different authors and, equally importantly, different translators. We excluded translations that are too free and too literary, such as (2), where the English version is a paraphrase relying on world knowledge rather than a translation.[5] In cases like these, the quality criteria of grammatical correctness, congruence in terms of content and comprehensibility (see Sect. 2) are obviously exceeded by stylistic considerations.

(2) Po levuju ruku za volnistymi zelenovatymi steklami *serebrilis'* groby poxoronnogo bjuro "Nimfa". (RNC: I. A. Il'f, E. P. Petrov. *Dvenadcat' stul'ev*. 1927)
Lit. 'On the left hand through undulating green glasses *shimmer.like.silver*$^{\text{PAST.ipf}}$ coffins of the funeral home "Nymph".'
'On the left you *could see* the coffins of the Nymph Funeral Home *glittering with silver* through undulating green-glass panes.'
(RNC: Ilya Ilf, Evgeny Petrov. *The Twelve Chairs*—John Richardson. 1961)

The resulting corpus thus encompasses 200 verbs in the past tense, indicative, active form, one half each in the ipf aspect and one in the pf aspect, for which a mapping between Russian and English is possible. An overview of the English correspondences to the Russian aspect forms as chosen by human translators is given in Sect. 3.3.

3.2.1 Translation of imperfective verbs

Out of 100 Russian ipf verbs 77 were translated into English by means of the simple past, cf. (3):

(3) On *stojal* na beregu ruč'ja.
(RNC: N. N. Nosov. *Priključenija Neznajki i ego druzej*. 1953–1954)
'It *stood* on the bank of a little brook.'
(RNC: Nikolay Nosov. *The Adventures of Dunno and his Friends*—Margaret Wettlin. 1980)

Six Russian ipf verbs were translated into English with the past progressive. As a rule, the specific interpretation underlying this translation needs some contextual trigger (this illustrates the non-equivalence of the aspectual systems of both languages; see also Sect. 3.3). This can be seen in (4) where the lexical meaning of the verb and the context—here the pf

[4] http://www.ruscorpora.ru/search-para-en.html.

[5] By referring to the shimmering as such and by referring to the visual perception of something shimmering, respectively, the Russian original and the English translation deliver different descriptions of reality.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 7/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

aspect *prišla*[PAST.pf] '[she] come' which serves as a reference point for the process described by ipf aspect *smotrela*[PAST.ipf] '[she] watch'—suggest a processual interpretation of the ipf aspect:

(4) Ona prišla k poezdu ran'še ego i *smotrela* na sxodjaščix vniz ljudej.
(RNC: N. A. Ostrovskij. *Kak zakaljalas' stal'*. Č. 2. 1930–1934)
'She had reached the station before him and *was watching* the people coming off the bridge.'
(RNC: Nikolai Ostrovsky. *How the Steel was Tempered*. Pt 2—R. Prokofieva. 1952)

In five cases, the temporal context obviously suggested the translation of Russian ipf aspect into English past perfect, cf. (5) where the conative interpretation of *ne ponimal*[PAST.ipf] '[he] not understand' triggers this particular translation for *videl*[PAST.ipf] '[he] see' and *ožidali*[PAST.ipf] '[they] wait':

(5) Nikto ne ponimal, počemu Pavku Korčagina vygnali iz školy. Tol'ko Serežka Bruzžak, drug i prijatel' Pavki, *videl*, kak Pavka nasypal popu v pasxal'noe testo gorst' maxry tam, na kuxne, gde *ožidali* popa šestero neuspevajuščix učenikov. Im prišlos' otvečat' uroki uže na kvartire u popa.
(RNC: N. A. Ostrovskij. *Kak zakaljalas' stal'*. Č. 1. 1930–1934)
'None of the children could understand why Pavel Korchagin had been ejected, none but Sergei Bruzzhak, who was Pavel's closest friend. He *had seen* him sprinkle a fistful of home-grown tobacco into the Easter cake dough in the priest's kitchen where six backward pupils *had waited* for the priest to come and hear them repeat their lesson.'
(RNC: Nikolai Ostrovsky. *How the Steel was Tempered*. Pt 1—R. Prokofieva. 1952)

In (5), the context is given by the preceding sentence, which includes a verb in the simple past: "None of the children *could* understand [...]"; this serves as a reference point describing a state following the situation in consideration, i.e. the situation described by *videl* and *ožidali*. Therefore, these verbs exhibit a pluperfect reading for the ipf aspect. This kind of interpretation is not yet considered in our groups of interpretation. It is based on the same constellation that holds for group III_{ipf} readings, i.e. the expression of an event in its entirety, seen, however, not from a contemporary point of view but from an anterior point of view.[6]

In three other cases the Russian ipf aspect is translated into the English present perfect; in all three examples the Russian verb is *byt'* 'to be', cf. (6) for an illustration:

(6) Učeba s nim *byla* nelegkaja.
(RNC: N. A. Ostrovskij. *Kak zakaljalas' stal'*. Č. 2. 1930–1934)
'Instructing him *has* not *been* easy.'
(RNC: Nikolai Ostrovsky. *How the Steel was Tempered*. Pt 2—R. Prokofieva. 1952)

In (6), the trigger is provided by the larger context which suggests a reading of current relevance of the ipf aspect. This interpretation again is not yet considered in any of our groups and, thus, has to be examined further.

The remaining nine examples are quite particular translations which can be explained by language-internal (English) rules[7] (see also (17) and (20) in Sect. 4.2): three of these trans-

[6] This example thus illustrates the 'secondary deictic' nature of aspect pointed out by Padučeva (2006).

[7] These nine translations are different from example (2), which exceeds by far language-internal rules and is a free paraphrase.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 8/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

lations are given by the additional modal verb 'would' [+ infinitive] (all in one sentence), cf. (7):

(7) Golovy rabočix *podnimalis'* vverx, glaza zadumčivo *tonuli* v serovatoj mgle, obnjavšej gorod, i často topor, zanesennyj dlja udara, nerešitel'no, na sekundu *ostanavlivalsja* v vozduxe, točno bojas' razrubit' laskovyj zvon.
(RNC: M. Gor'kij. *Ledoxod*. 1912–1915)
'[…] and at intervals heads *would raise themselves*, and blue eyes *would gleam* thoughtfully through the same grey fog in which the town lay enveloped, and an axe uplifted *would hover* a moment in the air as though fearing with its descent to cleave the luscious flood of sound.'
(RNC: Maxime Gorky. *The Icebreaker*—D. J. Hogarth. 1921)

One Russian verb form each is translated into English with an additional 'could' [+ infinitive], 'seemed to' [+ infinitive], 'used to' [+ infinitive], 'had come to' [+ infinitive], 'were becoming' [+ past participle] and 'kept' [+ present participle].

3.2.2 Translation of perfective verbs

The majority of Russian pf verb forms, namely 73 out of 100, were translated as English verbs in the simple past, cf. (8):

(8) Čestno govorja, prežde vsego ja *podumal*, čto èto utka.
(RNC: A. N. Strugackij, B. N. Strugackij. *Piknik na obočine*. 1971)
'To tell the truth, I first *thought* it was a hoax.'
(RNC: Arkady Strugatsky, Boris Strugatsky. *Roadside Picnic*—Antonina W. Bouis. 1977)

Ten Russian pf verbs were translated with English verbs in the present perfect, cf. (9), where a current relevance reading is suggested:

(9) Polnoč'. Uže davno *provolok* svoe razbitoe tulovišče poslednij tramvaj.
(RNC: N. A. Ostrovskij. *Kak zakaljalas' stal'*. Č. 2. 1930–1934)
'Midnight. The last tramcar *has* long since *dragged* its battered carcass back to the depot.'
(RNC: Nikolai Ostrovsky. *How the Steel was Tempered*. Pt 2—R. Prokofieva. 1952)

The English present tense is used as the historical present (narrative present) for Russian pf verbs in seven cases, cf. (10). This can again be explained by language-internal rules or conventions that apply only after the translation process:

(10) Luna *zalila* neživym svetom podokonnik. Golubovatym pokryvalom *leg* luč ee na krovat', otdavaja polut'me ostal'nuju čast' komnaty. Луна *залила* неживым светом подоконник.
(RNC: N. A. Ostrovskij. *Kak zakaljalas' stal'*. Č. 2. 1930–1934)
'The moon *lays* its cold light on the windowsill and *spreads* a luminous coverlet on the bed, leaving the rest of the room in semi-darkness.'
(RNC: Nikolai Ostrovsky. *How the Steel was Tempered*. Pt 2—R. Prokofieva. 1952)

In six cases the English past perfect is used, as in (11):

(11) Prozvali ego Gorškom za to, čto mat' *poslala* ego snesti goršok moloka d'jakonice, a on *spotknulsja* i *razbil* goršok. (RNC: L. N. Tolstoj. *Aleša Goršok*. 1905)

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 9/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

'He was called the Pot, because his mother *had* once *sent* him with a pot of milk to the deacon's wife, and he *had stumbled* against something and *broken* it.'
(RNC: Leo Tolstoy. *Alyosha the Pot*—Louise and Aylmer Maude. 1911)

The context is given here by the main sentence with a verb in the passive with the simple past ("He *was* called the Pot [...]"); this describes a state for which the subsequently described events provide an explanation. Therefore, the pf verbs receive a pluperfect reading.[8]

The remaining four examples are quite particular translations which again can be explained by language-internal rules and hence are not relevant for our purposes: one Russian verb form each is translated into English with the additional modal verb 'would' [+ infinitive], with 'came' [+ present participle], with 'had gone' [+ simple past] and 'happened to be' [+ adjective], cf. (12) for an illustration:

(12) Esli kakaja-nibud' malyška vstrečala na ulice malyša, to, zavidev ego izdali, sejčas že *perexodila* na druguju storonu ulicy.
(RNC: N. N. Nosov. *Priključenija Neznajki i ego druzej*. 1953–1954)
'If a girl-Mite caught sight of a boy-Mite coming down the street, she *would cross* to the other side.'
(RNC: Nikolay Nosov. *The Adventures of Dunno and his Friends*—Margaret Wettlin. 1980)]

3.3 Evaluation of translations

The study presented in Sect. 3.2 shows that the majority, i.e. three-quarters, of Russian past tense verbs (both ipf and pf) in the sample are translated by English simple past forms. This follows on in straightforward fashion from two facts: the markedness relations holding within the aspectual oppositions in both languages and the direction of translation. In Russian, the ipf aspect is the semantically unmarked member, being neutral as concerns the feature of 'completion' (which is expressed by the pf aspect), in English the simple form is semantically unmarked, being neutral with respect to the feature of 'continuous process' (which is expressed by the progressive form). As a consequence, the English simple form covers many uses of the Russian pf aspect and specific uses of the ipf aspect, while the English progressive form is an equivalent only to the continuous and the progressive reading of the ipf aspect, see Table 2 (cf. also Zangenfeind and Sonnenhauser 2014, p. 746).

Statistically, it is thus to be expected that Russian aspect can be translated into English by the simple form in most cases. And indeed, this is confirmed by the test corpus, see the summaries of results in Table 3.

There is no translation of the Russian ipf aspect that uses the English present tense and no translation of the Russian pf aspect that uses the English past progressive, which comes as no surprise. Moreover, there are five translations of the ipf aspect with past perfect (cf. example (5)) and three translations with the present perfect (cf. example (6)). These possibilities are not (yet) covered by the groups of interpretations listed in Table 3 and thus indicate the necessity of further refining the theoretical assumptions.

Given that 75 % of all Russian verb forms in the past tense are translated into English simple past, this study might be interpreted as suggesting that the problem of aspect interpretation constitutes a minor problem in the practice of MT, because verb forms other than those in the simple past are a minority. As a consequence, one might also argue that in many cases SMT produces correct results by using 'default translations' according to high proba-

[8]The question as to the difference between ipf and pf present perfect and pluperfect readings is beyond the scope of the present paper.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 10/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

Table 2 Markedness relations for aspect in Russian and English

Russian	English
pf aspect	simple form
ipf aspect	progressive form

Table 3 Overview of translations and aspect groups

Translations of Russian ipf in test corpus		Translations of Russian pf in test corpus	
Translation type	# of occurrences	Translation type	# of occurrences
simple past (group II/III$_{ipf}$)	77	simple past (group I$_{pf}$)	73
past progressive (group I$_{ipf}$)	6	past progressive	0
past perfect (no group so far)	5	past perfect (group III$_{pf}$)	6
present perfect (no group so far)	3	present perfect (group II$_{pf}$)	10
present tense	0	present tense (no group so far)	7
'would' [+ inf]	3	'would' [+ inf]	1
'could' [+ inf]	1	'came' [+ present participle (-ing)]	1
'seemed to' [+ inf]	1	'had gone' [+ simple past]	1
'used to' [+ inf]	1	'happened to be' [+ Adj.]	1
'had come to' [+ inf]	1		
'were becoming' [+ past participle]	1		
'kept' [+ present participle]	1		

bility, i.e. simple past in this particular case. And indeed, this is quite true for our test corpus: 84 of all 100 ipf verb forms and 87 of all 100 pf verb forms in the Russian sentences are translated using the English simple past by the statistical system Google translate.[9] This also means that the overall score in evaluation of Russian-English MT quality is influenced only to a small degree[10] by wrong translations of Russian aspect—at least if we assume that the high average of 75 % of English verb forms in the past are simple past also in other corpora as well. This means that if MT would translate every verb using the 'default' simple past, 75 % would be right anyway, which would be a quite good result for MT—compared to 72 % fluency and 67 % adequacy for the best SMT system described in Graham et al. (2014)—and which could be achieved without any effort.

Even relatively high success rates for the translation of aspect in SMT thus come as no surprise. They do, however, not tell us much about the rules of the aspect system in a language. And the 25 % of translations that do not use the simple past are not a small number. Moreover, and even more importantly, the cases in which Russian ipf verbs are marked by aspect-triggers are highly interesting for linguistics and the understanding of language.

[9] Obviously, not all of these SMT translations using simple past were in accordance with the translations by human translators in ruscorpora. Only 64 out of the 84 translations of ipf verb forms were translated in accordance with ruscorpora (cf. the 77 translations with simple past in ruscorpora); i.e. in 20 cases Google used simple past where human translators did not, and in 13 cases Google did not use simple past where human translators did. Concerning translations of pf verbs, the figures are: 66 out of the 87 Google translations were in accordance with ruscorpora (cf. the 73 translations with simple past in ruscorpora); i.e. in 21 cases Google used simple past where human translators did not, and in 7 cases Google did not use simple past where human translators did. The correctness of Google translations that are not in accordance with translations by humans must still be evaluated by native speakers.

[10] As an average just about 25 % of translations should be concerned.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 11/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

4 Implementation of rules

As pointed out in Sect. 2, RBMT offers the unique possibility to verify and improve theoretical assumptions by testing them in practice, i.e. by taking them as basis for the formulation of rules which are then implemented in a MT system. This will be shown in the present section on the example of two rules for aspect translation which we derived from theoretical considerations and successfully implemented in a RBMT system.

The theoretical assumptions have been described in Sect. 3.1. For high-quality MT the rules derived from these assumptions must cover—amongst others—all cases described in Sects. 3.2 and 3.3. In order to show the possibility of realizing aspect interpretation and translation rules in practice, we implemented, as a first step, some examples of such rules, two of which we presented in Sonnenhauser and Zangenfeind (2013), and Zangenfeind and Sonnenhauser (2014). For the implementation, the linguistic processor ÈTAP-3 (for an earlier version cf. Apresjan et al. 1989) was chosen, which translates from Russian to English and *vice versa* using dependency trees as an abstract level for transfer. ÈTAP-3 is a RBMT system that is quite elaborate but which, to date, processes aspectual information only rudimentarily. However, with its syntactic and semantic features[11] it provides an excellent basis for implementing rules for aspect interpretation and translation.

As has been pointed out in Sect. 3.1, RBMT of aspect involves two steps: language-internal interpretation, i.e. the assigning of a particular aspectual form to one of the groups listed in Table 1, and translation proper, i.e. the assigning of an appropriate form in the target language. That is, two sub-rules need to be formulated for every group, as will be illustrated in the following using two test cases as examples.

4.1 Test case 1

One of the rules implemented describes the interpretation of sentences that include an ipf unidirectional Russian verb of motion in the past tense and an adverbial that describes a situation in terms of its quality, as *medlenno* 'slowly', cf. (13):

(13) On polz medlenno, kak gusenica [. . .].
(RNC: M. A. Bulgakov. *Master i Margarita*. 1929–1940)
Lit. 'He creep$^{\text{PAST.ipf}}$ slowly like caterpillar.'

In this sentence the following information is relevant for MT:

(i) The ipf verb *polzti* '[to] creep' as a unidirectional verb of motion allows for the actual-processual interpretation;[12] for our purposes it is labelled with the semantic feature 'directional motion'.
(ii) The adverbial *medlenno* 'slowly' describes a situation in qualitative terms; thus, it is labelled with the syntactic feature 'quality'.[13]

The corresponding sub-rule 1, which applies for language-internal interpretation, states that the verb in a sentence displaying these conditions is assigned to an interpretation of group I_{ipf} (see Table 1) and has thus to be labelled with the feature 'group I_{ipf} interpretation'. In a more technical form this rule is shown in (14):

[11] The differentiation of syntactic and semantic features in ÈTAP is mainly due to technical reasons.

[12] Cf. Table 1. For a classification of predicates cf. Apresjan (2006); his classification includes 17 classes, some of which exclude certain disambiguation possibilities and/or make others highly probable.

[13] This syntactic feature had to be introduced to ÈTAP because a similar already existing feature could not be used for our purposes.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 12/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

(14) IF
X is a finite verb
AND IF
X has the semantic feature 'directional motion'
OR
X is a support verb AND X governs a noun Z1 AND Z1 has the semantic feature 'directional motion'[14]
AND IF
X governs an adverbial Z2 AND Z2 has the syntactic feature 'quality'
AND IF
X has the morphological feature 'past tense' AND has the morphological feature 'ipf'
THEN
X is labelled with the feature 'group I_{ipf} interpretation'

Sub-rule 2, i.e. the rule applying for translation, is a simple instruction stating that the morphological feature 'past tense' of Russian verb forms that have been assigned to the feature 'group I_{ipf} interpretation' is replaced by the morphological feature 'past progressive' for the corresponding English verb form, cf. (15):

(15) IF
X has the feature 'group I_{ipf} interpretation'
THEN
the morphological feature 'past tense' of X is replaced by the morphological feature 'past progressive'

Applying these two rules, the sentence in (13) is translated by ÈTAP as in (16):

(16) He was creeping slowly as a caterpillar.

The translation by human translators—with the synonymous verb *crawl*—is quite similar:

(16′) It was crawling slowly along like a caterpillar.
(RNC: Mikhail Bulgakov. *Master and Margarita*—Richard Pevear, Larissa Volokhonsky. 1979)

SMT Google translate (1 March 2016) produces the following sentence using the 'default' simple past instead of the past progressive:

(16″) He crawled slowly, like a caterpillar.

4.2 Test case 2

Another rule that has already been implemented describes the interpretation of sentences that include an ipf Russian verb, a nominal predicate of the class 'occupation' and an adverbial expressing regularity, as in (17):

(17) Ran'še ja po večeram prodelyval ėti gimnastičeskie upražnenija.[15]
Lit. 'Formerly I in evenings do$^{PAST.ipf}$ these gymnastic exercises.'

[14] This line is necessary for sentences in which the predicate is given by a support verb construction, cf. (17) below.

[15] This is a simplified example from Bendixen et al. (2005–2012).

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 13/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

The sentence in (17) involves various kinds of aspectually relevant information; for the purpose of MT, three facts are decisive:

(i) The adverbial phrase *po večeram* 'in the evenings' expresses regularity; thus, it excludes group I_{ipf} and group III_{ipf} interpretations (cf. footnote 10). The attribute of regularity is technically deduced as follows: the preposition in this adverbial phrase is *po1* in ĖTAP's dictionary, i.e. one of two lexemes of the preposition *po* that govern the dative case in ĖTAP's dictionary.[16] For our purposes, this preposition has to be differentiated further: the lexeme employed in sentence (17) governs a word form of a temporal lexeme and corresponds to *po17* of the Slovar' russkogo jazyka (1983).[17] Because we did not want to change the system of lexemes in the dictionary of ĖTAP, the differentiation is carried out by a part of our rule for language-internal aspect interpretation, cf. (18) below. This part of the rule in question simply checks whether *po1* of ĖTAP governs a lexeme that has the semantic feature 'time' and the syntactic feature 'duration' (i.e. a temporal lexeme) and thus is mapped to *po17* of the Slovar' russkogo jazyka (1983). In the lexical entry of *po1* in ĖTAP it is enough to label it with the syntactic feature 'regularity'.

(ii) The verb *prodelyvat'* '[to] do' is used as a support verb in the context of the given sentence; i.e. it has no lexical semantics (cf. Mel'čuk 2004) but merely contributes aspectual information (= ipf). This information is provided by the morphological dictionary of ĖTAP.

(iii) The noun *upražnenie* 'exercise' is the semantic predicate in the sentence. It belongs to the semantic class *zanjatie* 'occupation', according to Apresjan (2006, p. 83, 86f). Thus, it is labelled with the semantic feature 'occupation'. In combination with an ipf support verb such as *prodelyvat'* it allows for all groups of ipf interpretations.

The rule applied for language-internal aspect interpretation of sentences like in (17) works as follows:

(18) IF
X is a finite verb
AND IF
X has the semantic feature 'occupation'
OR
X is a support verb AND X governs a noun Z1 AND Z1 has the semantic feature 'occupation'[18]
AND IF
X governs an adverbial expression whose head is Z2 AND Z2 is a preposition
AND
Z2 has the syntactic feature 'regularity' AND Z2 governs a lexeme Z3 AND
Z3 has the semantic feature 'time' and the syntactic feature 'duration'[19]

[16] Two further lexemes of the preposition *po* in ĖTAP govern the accusative and the prepositional case, resp.

[17] This dictionary lists 20 lexemes of the preposition *po* that govern the dative case. In the course of developing the rules presented in this paper *po17* proved to be the lexeme that fits best here, other than assumed in Sonnenhauser and Zangenfeind (2013) and Zangenfeind and Sonnenhauser (2014), where *po16* was used.

[18] The sentence in (17) is an example where this line is necessary because of the support verb construction.

[19] This line checks e.g.—if the preposition in question is *po1* of ĖTAP—whether the preposition *po* corresponds to *po17* of the Slovar' russkogo jazyka (1983). For further cases adverbial expressions in the form of simple adverbs with the syntactic feature 'regularity' would have to be also considered. For simplicity's sake this has not been done yet.

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 14/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

AND IF
X has the morphological feature 'past tense' AND has the morphological feature 'ipf'
THEN
X is labelled with the feature 'group II_{ipf} interpretation'

For pure translation again just a very simple rule has to be applied:

(19) IF
X has the feature 'group II_{ipf} interpretation'
THEN
the morphological feature 'past tense' of X is replaced by the morphological feature 'simple past'

The result is the sentence in (20), in which only the word order must still be arranged:

(20) Earlier I in the evenings did these gymnastic exercises.

The most adequate translation, i.e. the English habitual construction 'used to', can be described by means of language-internal rules (cf. example (7) above and further remarks there) and is not necessarily an immediate concern of translation.

In this case Google translate (1 March 2016) uses the past progressive:

(20′) Earlier in the evening I was doing the gymnastic exercises.

The reason for why Google translate does not use the 'default' option simple past in this context is not clear. In any case this example shows once more that linguistically based rules are indeed necessary for aspect translation.

The rules presented in this chapter illustrate the basic feasibility of our approach. Of course, further rules need to be identified, described and implemented, but a start has been made.

5 Conclusion

The example of Russian to English aspect translation shows that there are domains for which RBMT is not only beneficial but even necessary. Well-formulated rules allow pure chance in translation to be replaced by regularity; moreover, by this well-formedness criterion theoretical linguistic assumptions can be verified or falsified and hence be improved. This has been illustrated in the present paper by a sample of 200 sentences taken from the Russian-English parallel corpus of the Russian National corpus. This sample showed that about 75 % of Russian verbs in the past tense are translated into the English simple past. Nine percent of ipf verbs and four percent of pf verbs had quite specific translations that were not further considered for our purposes. Taking the simple past as a default translation for both aspects would thus result in an acceptable success rate. This success rate is indeed observed for statistical machine translation.

However, the cases that are not captured by this default assignment are by no means negligible—neither from a theoretical nor a computational point of view. Of special interest are the 14 % of ipf verbs that are translated in the sample using the English past progressive (belonging to 'group I_{ipf}') as well as the past perfect and present perfect, which do not belong to any of the semantic groups that have been posited from a theoretical perspective. These cases necessitate an improvement of the underlying theoretical assumptions and illustrate the

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF

« RUSS 11185 layout: Small Condensed v.2.1 file: russ9169.tex (Loreta) class: spr-small-v1.2 v.2016/06/09 Prn:2016/09/02; 14:18 p. 15/15»
« doctopic: OriginalPaper numbering style: ContentOnly reference style: apa»

inadequacy of default mappings in translation. The same holds for the 16 % of pf verbs that are translated into the English past perfect and present perfect. These cases, too, are insightful for the analysis of Russian aspect as they also belong to those cases for which 'default' translation does not work.

As it includes only 200 examined verbs and is restricted to Russian literature dating from the 20th century, the corpus for this study is not representative. Nevertheless a tendency can be seen to exist.

In a next step, a larger corpus will be examined in order to further analyse the mechanisms of aspect interpretation. Based on this analysis, further rules for the disambiguation (step 1) and translation (step 2) of aspect will be formulated and implemented in ÉTAP-3.

References

Apresjan, Ju. D. (2006). Fundamental'naja klassifikacija predikatov. In Ju. D. Apresjan (Ed.), *Jazykovaja kartina mira i sistemnaja leksikografija* (pp. 75–110). Moskva.

Apresjan, Ju. D., et al. (1989). *Lingvističeskoe obespečenie sistemy ÉTAP-2*. Moskva.

Bendixen, B., et al. (2005–2012). *Russisch aktuell*. Wiesbaden.

Bickel, B. (1996). *Aspect, mood, and time in Belhare: studies in the semantics-pragmantics interface of a Himalayan language*. Zürich.

Dobrovol'skij, D. O., Kretov, A. A., & Šarov, S. A. (2005). Korpus parallel'nyx tekstov: arxitektura i vozmožnosti ispol'zovanija. In *Nacional'nyj korpus russkogo jazyka: 2003–2005* (pp. 263–296). Moskva. Retrieved from http://ruscorpora.ru/sbornik2005/17dobrovolsky.pdf (1 March 2016).

Graham, Y., Baldwin, T., Moffat, A., & Zobel, J. (2014). Is machine translation getting better over time? In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL-14). Gothenburg, Sweden, April 26–30, 2014* (pp. 443–451). Gothenburg.

Iomdin, L. L. (2003). Purpose and idea: a lesson drawn from machine translation. In *MTT 2003. First International Conference on Meaning-Text Theory. Paris, Ecole Normale Superieure* (pp. 269–278). Paris.

Iomdin, L. (2008). A few lessons learned from rule-based machine translation. In G. Gross & K. U. Schulz (Eds.), *Linguistics, computer science and language processing. Festschrift for Franz Guenthner on the occasion of his 60th birthday* (Tributes, *6*, pp. 171–187). London.

Klein, W. (1995). A time-relational analysis of Russian aspect. *Language*, *71*(4), 669–695.

Kulagina, O. S. (1979). *Issledovanija po mašinnomu perevodu*. Moskva.

Mel'čuk, I. (2004). Verbes supports sans peine. *Lingvisticæ investigationes*, *27*(2), 203–217.

Padučeva, E. V. (1992). Towards the problem of translating grammatical meanings. *Meta: Journal des traducteurs*, *37*(1), 113–126.

Padučeva, E. V. (2006). Nabljudatel': tipologija i vozmožnye traktovki. In N. I. Laufer, A. S. Narin'jani, & V. P. Selegej (Eds.), *Komp'juternaja lingvistika i intellektual'nye texnologii: Trudy Meždunarodnoj konferencii "Dialog 2006". Bekasovo, 31 maja–4 ijunja 2006 g*. Moskva. Retrieved from http://www.dialog-21.ru/digests/dialog2006/materials/html/Paducheva.htm (1 March 2016).

Slovar' russkogo jazyka (1983): Evgen'eva, A. P. (Ed.) (1983). *Slovar' russkogo jazyka* (Vol. III). Moskva.

Sonnenhauser, B. (2006). *Yet there's method in it. Semantics, pragmatics, and the interpretation of the Russian imperfective aspect*. München.

Sonnenhauser, B., & Zangenfeind, R. (2013). Towards machine translation of Russian aspect. In V. Apresjan, B. Iomdin, & E. Ageeva (Eds.), *Proceedings of the 6th International Conference on Meaning-Text Theory. Prague, August 30–31, 2013* (pp. 192–201). Retrieved from http://meaningtext.net/mtt2013/proceedings_MTT13.pdf (1 March 2016).

Zangenfeind, R., & Sonnenhauser, B. (2014). Russian verbal aspect and machine translation. In V. P. Selegej et al. (Eds.), *Komp'juternaja lingvistika i intellektual'nye texnologii: Po materialam ežegodnoj Meždunarodnoj konferencii "Dialog". Bekasovo, 4–8 ijunja 2014 g. Vyp. 13*(20). Moskva. Retrieved from http://www.dialog-21.ru/digests/dialog2014/materials/pdf/ZangenfeindRSonnenhauserB.pdf (1 March 2016).

Journal ID: 11185, Article ID: 9169, Date: 2016-09-02, Proof No: 1, UNCORRECTED PROOF